CH5230: System Identification

Inputs for identification

Part 1

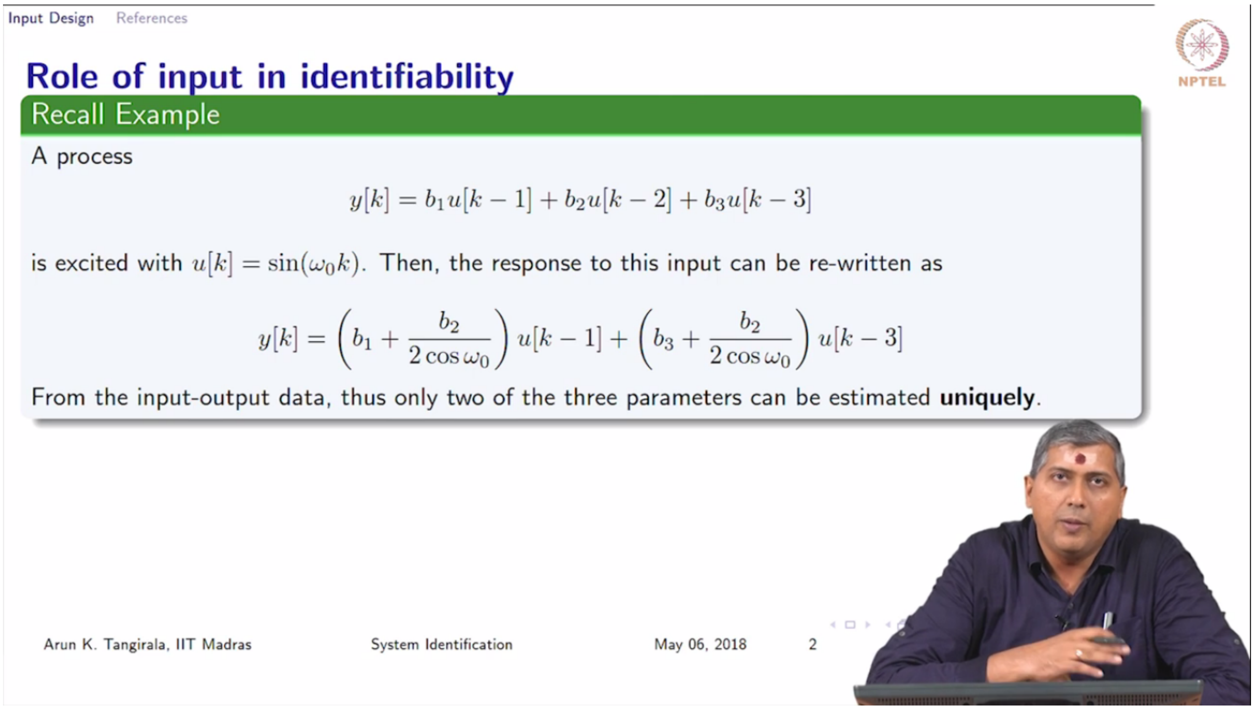
Okay. So welcome to the last lecture on input design. This is actually the last lecture in the topic, in the course on system identification. So all along in this course we have assumed, in fact from time to time I mentioned this that the input is persistently exciting. We did talk about the input briefly early on in the course when we talked of identifiability. So if you recall identifiability has three aspects to it. One has got to do with the nature of the model itself. That is, that there should be a unique model for a given input-output system. And secondly that, so second aspect of identifiability has got to do with input design and the signal to noise ratio. That you have asked enough questions and the interview process is a nice analogy to understand today's lecture. And we have used this analogy before as well.

So the second aspect of identifiability demands that enough questions and the right questions have been asked so as to identify the model uniquely. So the first requirement is that for identifiability, model should be unique. Once you are guaranteed there exists a unique model like a transfer function model without zero pole cancellation. Then you have asked enough questions, that means, you have excited the process well enough to obtain a unique model. And the third aspect of identifiability is the estimator itself. So asking enough questions, having a unique model alone, these two alone are not enough to be able to get a unique model.

And to be able to get to the truth, the estimator should also help you or help us reach that unique model. And that is where the consistency of the estimator plays a critical role. We have talked about consistency of the estimator. We have talked about the uniqueness of models. We have shown for example, states-space models are not unique, unless we impose some structural constraints on the state space mattresses and so on.

And now it is time to talk about input design. And ironically input design is a first step in identification and we are talking towards the end. And that's because a lot of the theory that goes into the input design, at least some of it will require the discussion that we have had on models and estimator and so on. And that's the reason we are pushed it to the end. So let's now get into, get onto this journey and ask, what is it that the theory tells us on how to design inputs for identification.

(Refer Slide Time: 2:59)



So before we jump into the theory, it may be a good idea to recall this example that we have had at the beginning of the course. Again when we talked of identifiability. In this example we have three coefficient FIR model with a unit delay. And the purpose of this example was and is to show that if the input is not sufficiently excited we may not be able to identify all the parameters uniquely. So here is a three parameter model and we excite the system with a sinusoidal input of some frequency, nonzero frequency.

And we have shown already this that. When you look at the response of the system to this input. Then you can you can rewrite the output as a two parameter model. Which means that I will not be able to identify the third parameter and whichever, maybe a third parameter. You can pick. You can say that I maybe, I will be able to identify B2 and B3 but not B1 or. In fact, we may be ending up identifying a linear combination of these three parameters. Two linear combinations as you can see. So in fact we'll be able to identify this coefficient and this coefficient but not necessarily the other coefficient the third one.

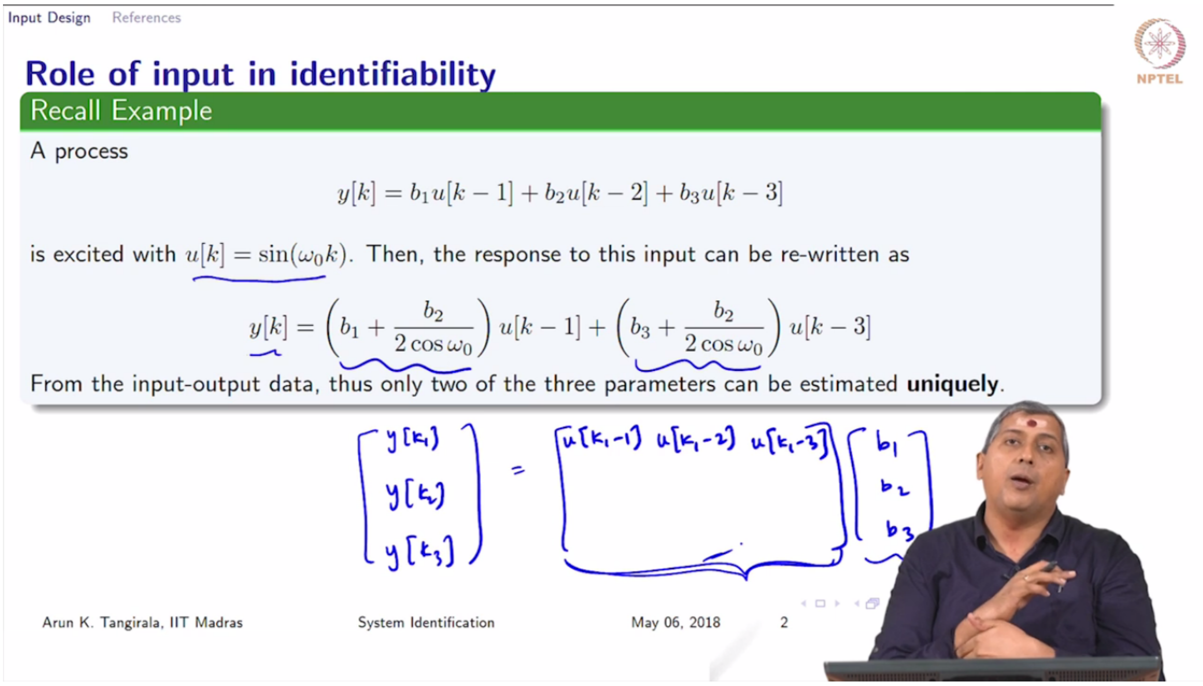
Of course the linear algebra perspective is what we have talked about as well. When you, this is a nice free system, linear system. So to estimate the three parameters you need to set up three equations. And when you write down the three questions at three different instance y k 1, y k 2 and y k 3, if we were to write down here the equations with B1 B2 B3 as the unknowns. It turns out. Let me just complete the first row here and then the remaining two rows follow suit. So I have u at k 1 minus 1, u at k 1 minus 2 and u at k 1 minus 3. And likewise for the remaining two rows.

So this matrix here plays a critical role. Simple linear algebra tells us that this matrix here plays a critical role in being able to estimate these three parameters. So it's important that this three by three matrix consisting of the inputs be a full rank. Right? Then only I can obtain a unique solution. It turns out that when u k is a sinusoid of a single frequency. This matrix here is rank deficient. So that's a linear algebra perspective of loss of identifiability. We have talked about that earlier.

Now it's important to remember this matrix. Later on we will replace this matrix with the covariance of the inputs rather than just the inputs itself. Because here there's no noise. We are directly working with the inputs. We're in the noisy conditions we work with covariances. So on the left hand side you would have cross-covariance of y with u. And on the right hand side, this big three by three matrix would be a variance covariance matrix of the inputs.

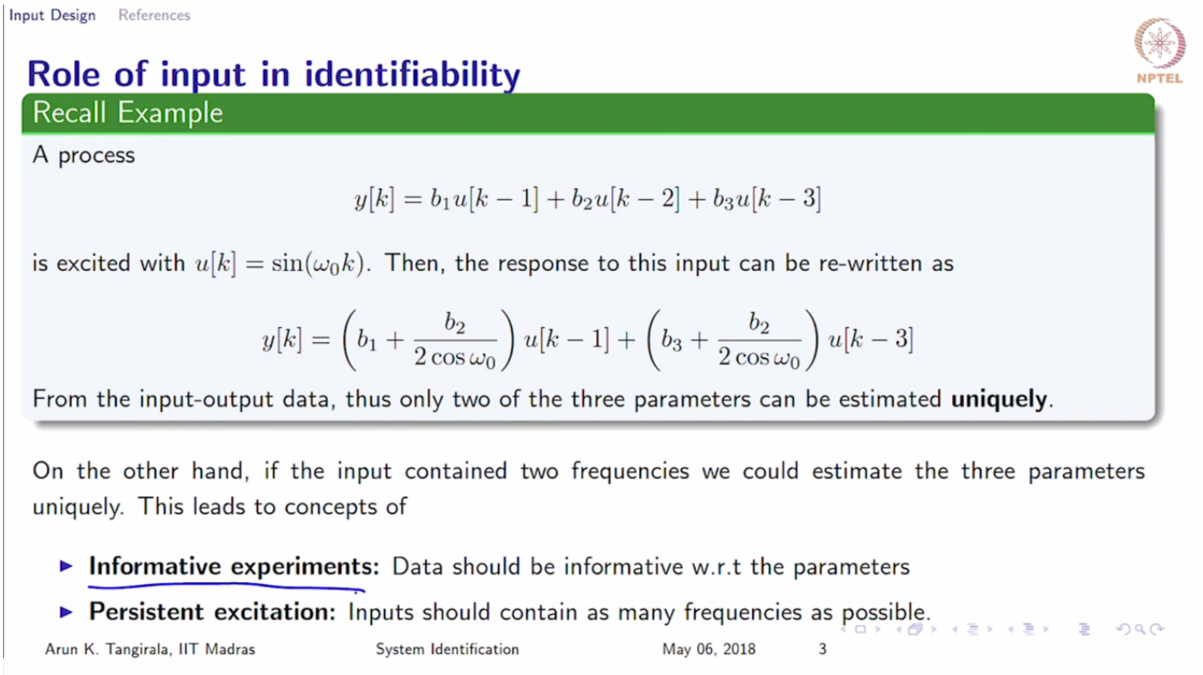
Because we set up the questions in the covariance domain. So there we will demand that the covariance of the inputs be full of full rank. Either way the simple message is that with a sinusoidal excitation of a single frequency I will not be able to estimate a three parameters model. In fact precisely I'll be able to estimator a two parameter model. So the rank of this matrix here is two, then u k is sinusoid of a single frequency. Of course if u k is a mix of two signs of different frequencies then this matrix becomes full rank. And you may be able to, you will be able to estimate up to four parameters.

(Refer Slide Time: 06:56)



So the important thing is that we want to generate informative. We want to perform informative experiments or we want to generate informative data. In the example that we just discussed there is insufficient information with a single frequency we are not generating enough information to estimate the three parameters. Now again the interview process or the analogy really helps. When I'm conducting an interview and I want to determine whether a particular candidate is suitable for the advertised position.

(Refer Slide Time: 07:31)

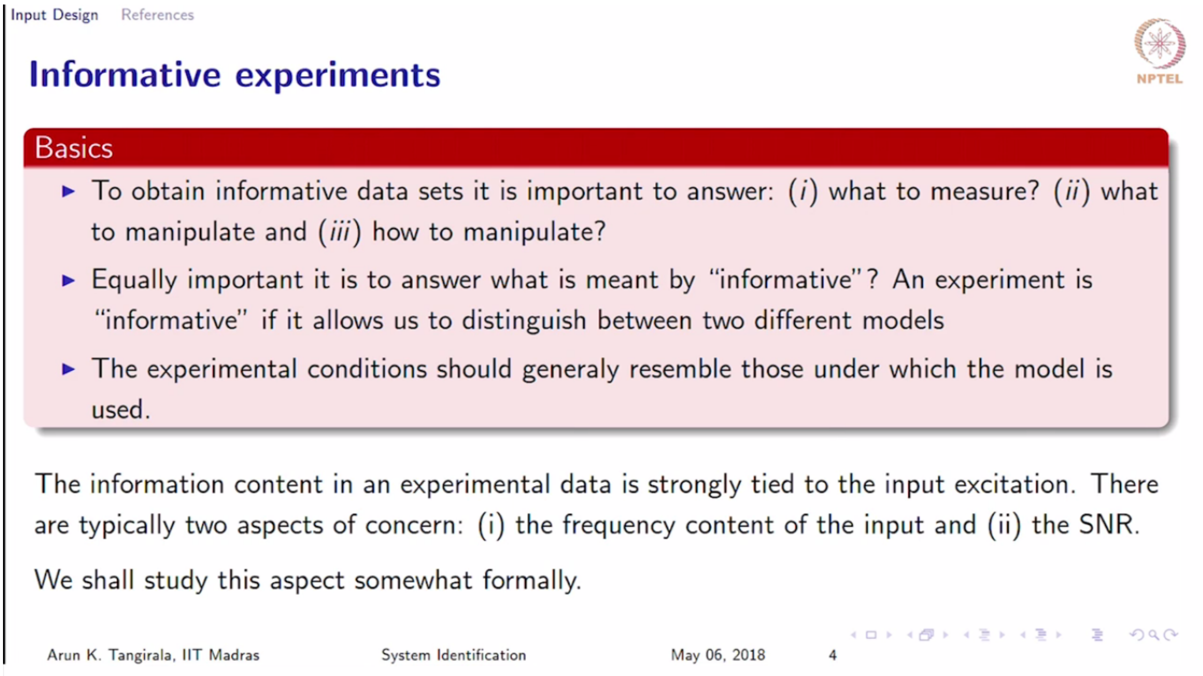


Then I should ask enough questions and I should ask right questions so that I can gather the information that is required for me as an interviewer, as an employer to make a decision. Right? And if I don't ask enough then I will be in a dilemma as to which candidate to choose because I have not asked enough questions to separate the most suitable from the less suitable. Secondly I should speak loud enough.

So if I were to ask my questions I may be asking the right questions and the right amount of questions also but if I don't speak loud enough. So if I say, you won't be able to the, candidate won't be able to hear. Or maybe the candidate may confuse the sound of the AC or the fan with my question and not able to answer clearly. Then again the information obtained is not good enough. So there is a signal to noise ratio also that plays an important role.

Now technically an identification, generating informative data is related to what is known as a persistent excitation, which means that, effectively it means that I have to fill the input with a lot of frequencies. We will learn the technical definition of persistent excitation in this lecture shortly.

(Refer Slide Time: 08:49)



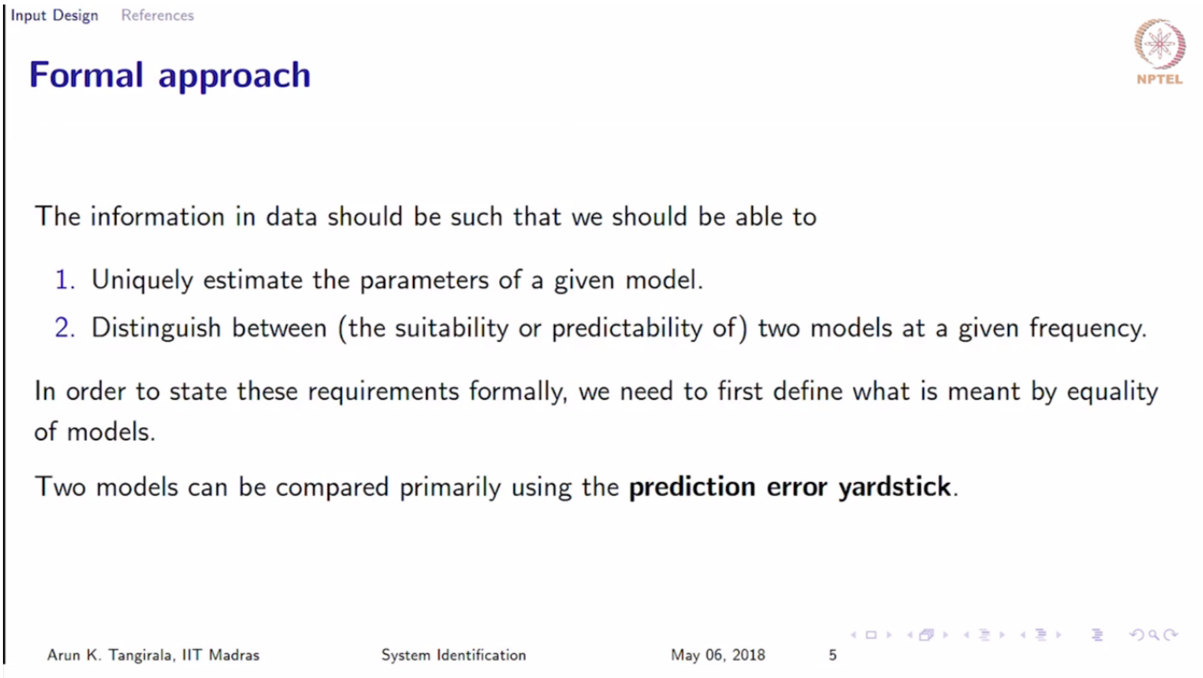
Okay. So just a few remarks on informative experiments. Basically to obtain informative data we want to ask what to measure. That means which variables to measure because I want to get the right information. And what to manipulate? Which means which input should I play around with and how to manipulate, which is input design itself. And what is important is this informative, the word information itself. We have talked about Fisher's information before.

Information in itself does not make much sense. Always information is in the context of some parameter estimation, some model estimation and so on. So I cannot ask the question, is the data informative? That's an incomplete or an ill post question to which there is no definitive answer. On the other hand it is okay to ask is the data informative with respect to this model or with respect to that parameter. Then it's a well [09:49 inaudible] question and we may be able to find answers. Right?

And that's the kind of questions that we are going to ask very soon. And we have already discussed the information content is strongly tied to the input excitation. That means how we are going to part of the system or question the system. And what we mean by input excitation is of course as we have already mentioned the frequency content and the signal to noise ratio.

So once again we step into the frequency domain world to answer the questions as much as some of us may be uncomfortable with the frequency domain. We have probably very little choice. These questions are better answered in the frequency domain. There are some answers in the covariance domain as well. We'll talk about that soon.

(Refer Slide Time: 10:35)



So what are the requirements for, now we are getting into some formalism here and therefore we want to ask the right questions and very questions with clarity? So what do we want? What do we mean by informative data and what is that kind of, what should this information help us? It should help us in uniquely estimating the parameters of a given model. There should be no ambiguity there.

And more or less it is also tied to be able to distinguish between the suitability or predictability of two models at a given frequency. What does this mean? So let's say I give you input-output data of system and you are guessing whether a first order or a second order is suitable. Right? It's a standard question. Or maybe with respect to the previous example I gave you input-output data and I told you the relationship between input and output is an FRI relationship.

Now you're debating whether it's a one parameter FIR model, two parameter FIR model, or a three parameter FIR model and so on. So let's say our model set is containing one, up to three parameter models. The previous example showed that if I [11:54 inaudible] the system with the sinusoidal input of a single frequency, then I will be able to distinguish between a one parameter and two parameter model. What do I mean distinguish, I will be able to answer whether a two parameter model is better suited or one parameter model is better suited.

On the other hand I will not be able to answer if a three parameter model is better suited because there is no unique estimate of the three parameter model with a sign of single frequency. We have already seen that which means we failed in our first objective. I may choose a three parameter model but the data does not have enough excitation or information to estimate a three parameter model uniquely.

And therefore I will not be able to meet objective two, which is to be able to answer whether a two or three parameter model is better suited to it. Both are tied to each other. You can see that. So in general you have seen and you have experienced many times even in your home problems that system identification at its heart is about selecting a model structure or choosing between two competing, two or more competing model structures.
And therefore it's very important that there is information in the data. To be able to say which model is better suited. Right? So now it all of it now boils down to, formerly being able to distinguish or resolve or discriminate between two model structures. Or you can say it is related to the equality of models. And essentially comparisons. So two models can be compared using the prediction error yardstick.

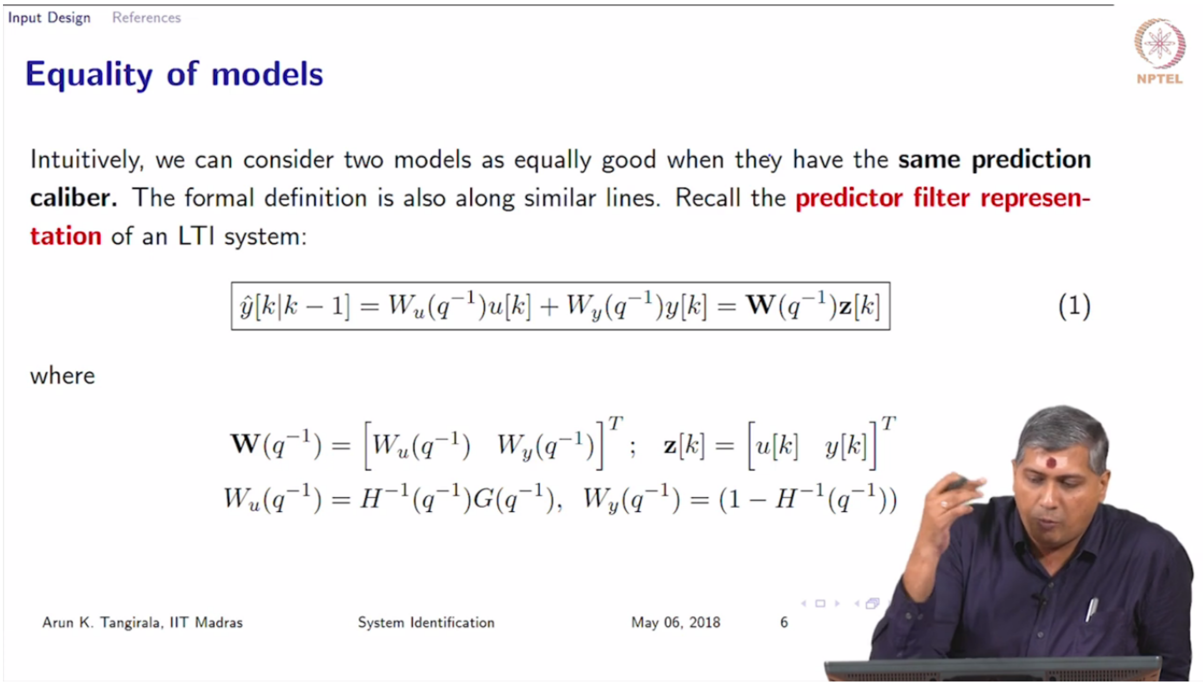
So how do I say two models are, one model is better suited than the other? After all why every building models? We had building models for predictions. So when I say the third order model is better than the second order model, let us say. It is almost implicit that the third order model is going to give me better predictions in the second model. So prediction error yardstick is a very natural way of comparing two models.

So the idea here is as follows. We are now going to hang on to the prediction error branch and then define equality of models. And then ask when do I fail to distinguish between two models. That is when meaning, not at a certain date or day or the time and so on. Under what conditions, what input excitation will fail me in distinguishing between two models. That means viewed through the lens of the input, If I end up seeing two models as being equal then-- Although there are of different structures then that's not such a great thing to happen.

I know that one is a higher order. Another is a lower order. But when viewed through the lens of the input, if these two models appear to be the same then that's not such a great thing to happen. What we want. Remember our input. Our questions are our windows of access to the process. So we want to ask the right questions, we want to open up the right windows so that we can see the process characteristics clearly and so that I can say without ambiguity that a lower order model or a higher order model is better suited than the other.

But if I don't ask enough questions, it may so happen and that's what the result is going to show that two models although mathematically are of different structures will appear the same. Pretty much like what you see in sampling. Right? In sampling what did we learn? We said that if we don't sample fast enough then a signal of very high frequency can manifest as a low frequency. And therefore will not be able to. I will not be able to distinguish and that is what we call as aliasing. So we know what we want to why that kind of aliasing here in input design.

(Refer Slide Time: 16:12)



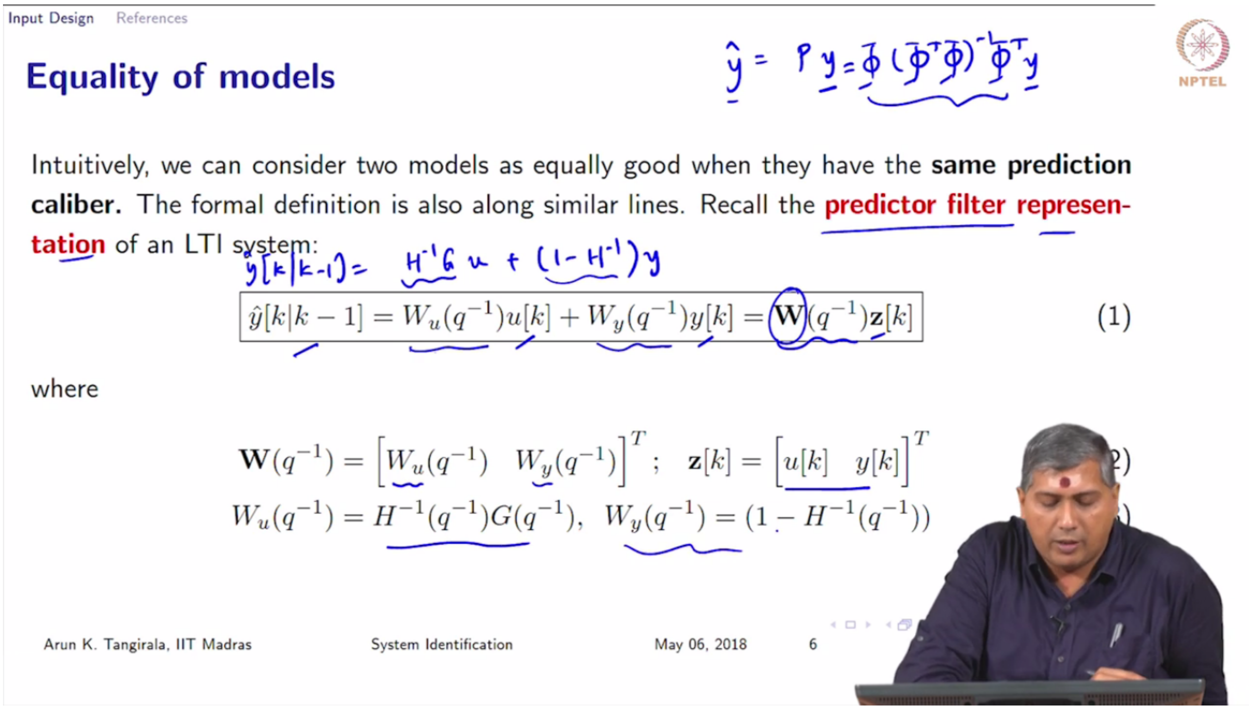
So we have talked about equality of models before briefly. In fact we introduced this idea of a predictor filter representation of an LTI system. You must recall this when we talked about predictions. So we said that. We could write the one-step-ahead prediction of an output as a linear combination of filtered input and output. So all along we have been using the symbols G and H the plant and noise models and we have shown also y hat of k, given k minus 1 is H inverse G u plus 1 minus H inverse y.

This is when we're given G and H. But we said effectively H inverse G is some kind of a filter W u and 1 minus H inverse years is another filter W y. So you can see the prediction essentially as a filtering of the input and the output. We see this even in linear regression. If you recall least squares, we said we could write the predictions as some projection matrix times y. So what is this projection matrix there, it is Phi transpose Phi inverse Phi transpose. Right? So this is our projection matrix.

How did we get this? Well we got this from our Phi theta. Right? If you recall. We had y hat equals Phi theta and theta hat is Phi transpose Phi inverse Phi transpose y. So the projection matrix is Phi times Phi transpose inverse times Phi inverse. So this is our projection matrix. You can see that I avoid the model estimation completely in order to construct the prediction. Likewise here, I don't have to estimator G and H, this equation if you recall, tells us that if I want to simply compute the prediction, one-step-ahead prediction I simply have to build optimal filters W u and W y. That is another way of looking at the prediction problem.

And we combine this input filter and output filter into a vector and that is our big W. You can say matrix of filters. And z is the [18:40 inaudible] input-output data. These are standard notation. Of course you could flip the order. You can define W as W y W u then you can write z as y and u. Okay. So I have already written the expression for W u and W y.

(Refer Slide Time: 18:51)



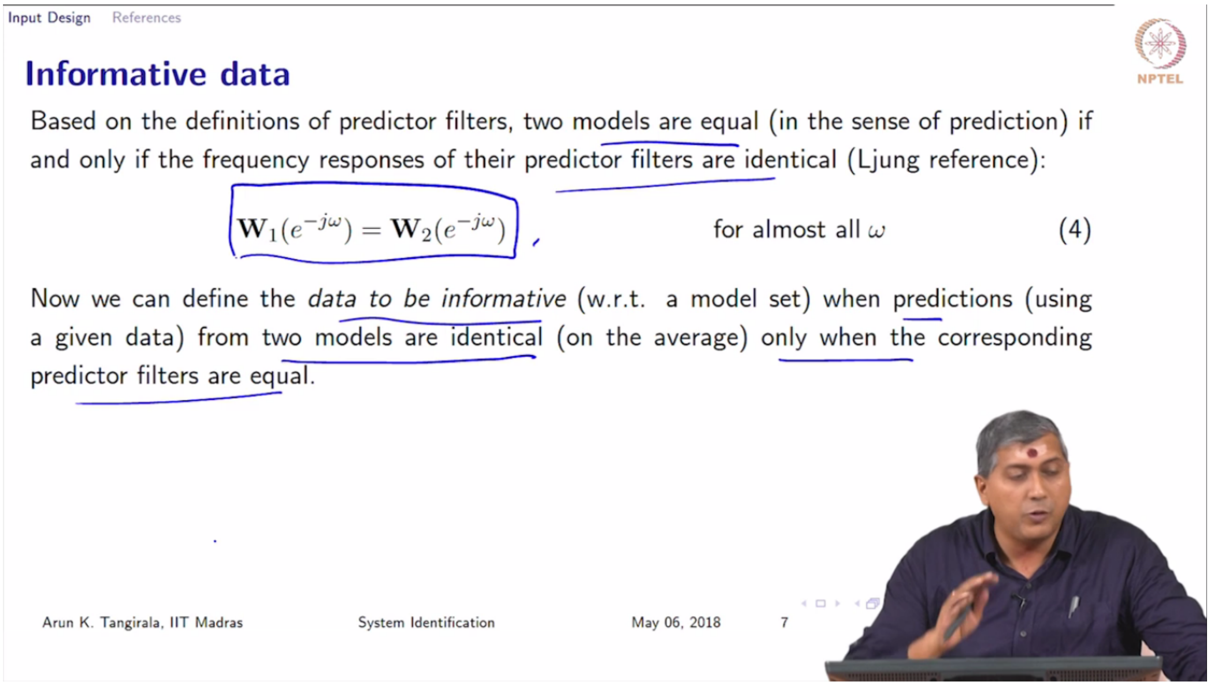
Okay. So how does this predictor filter approach help us in arriving at equality of models. So here is an important statement. We say that two models are equal. Now we're moving to a non-parametric domain. We would say that two models are equal when that predictor filters are equal. Because if their predictor will have equal the predictions are equal. And if predictions are equal for all practical purposes those two models are equal. Right? So the only difference is, we have written this in terms of the frequency response functions of the predictor filters but that doesn't matter. The reason for writing it that way is to make things non-parametric so that we don't think of parametric models all the time.

So remember this, we say that the two models are equal if their predictor filters are identical. Which is a nice framework, of course, you can. You're most welcome to define equality of models in a different way. But this is a very natural way of defining equality of models. Now we can define the data to be informative when predictions from two models are identical only when the corresponding predictor filters are equal. See, how our predictions going to be constructed from input and output data?

So now what we are going to say is, we are going to term or label data as informative only when, that is, when the predictions are equal if and only if when the two models are identical. If the two models are not identical my data should tell me clearly that they are not the same. How? Because the predictions will turn out to be different. Obviously.

See if two models are not the same. Typically predictions are going to be different but it is likely that the input and output does not have enough excitation in them that despite it using two different model structures I can get same predictions. That is exactly what I was mentioning earlier. That's kind of aliasing that you may end up having if your data doesn't have enough excitation. So with respect to the earlier example we had a three parameter FIR model. Right?

Whether I use a two parameter FIR model or a three parameter FIR model with the kind of excitation that they had I will get the same prediction that disappoint. Which means that the data has failed me in discriminating between two parameter and three parameter model.
(Refer Slide Time: 21:33)



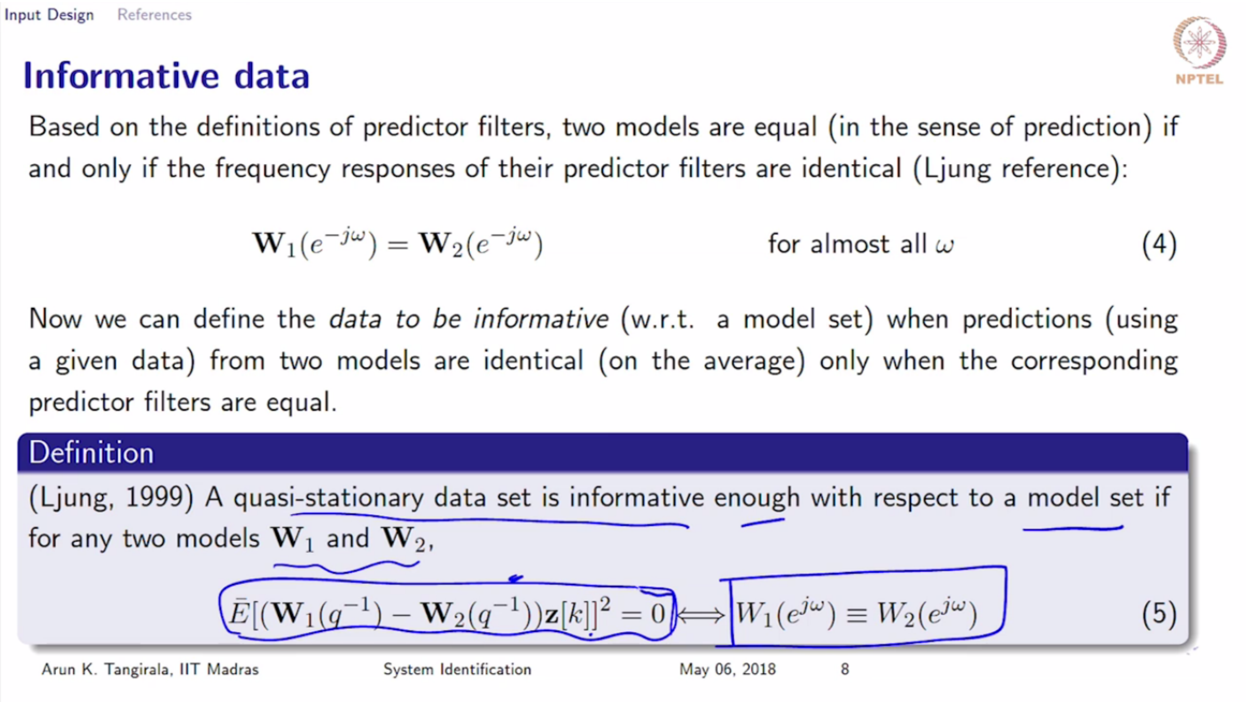
So formally now we see that the quasi-stationary data set is informative enough with respect to a model set. So your model set is your choice. Your model set may consist up to fourth order model or your model set may consists of all models. The choice is yours. Okay. So you pick a model set and with respect to that model set. Your data is informative.

Why do we say with respect to that model set? Because after all informative data has got to do with the ability to resolve or distinguish between two models and maybe there is just enough information to resolve models up to a certain order but beyond that there is no information. Right? So in the FIR example I am able to distinguish between one parameter and two parameter models. I can say clearly yes they are different with respect to the prediction. But if I include third order model also, three parameter models also, then the data that is generated by a single frequency excitation is not informative enough.

So formally we say that the data set is informative enough with respect to a model set, if or any two models W1 W2 in the model set, the average, the mean, that is the average squared error between the predictions is zero. Why did we bring expectation all of a sudden because we are going to deal with noisy data. So we can't really say that, "Oh, I will say W1 and W2 are equal when their predictions are identical. They may be identical for one realization, another realization they may not be.

S we want to be able to say this across all realisations of noise. What we are saying essentially is, if it so-- if it happens that the predictions from two different models are identical in a mean square sense it should only happen when the two models are identical. That is this should be true this boxed one should be true only when this is true. All right? If it means, if this is zero and this is not the reason for the predictions to be identical then that means the data set is not informative.

(Refer Slide Time: 24:10)



The interpretation is essentially if you leave aside all the technicalities a dataset is informative if it allows us to distinguish between two different models.

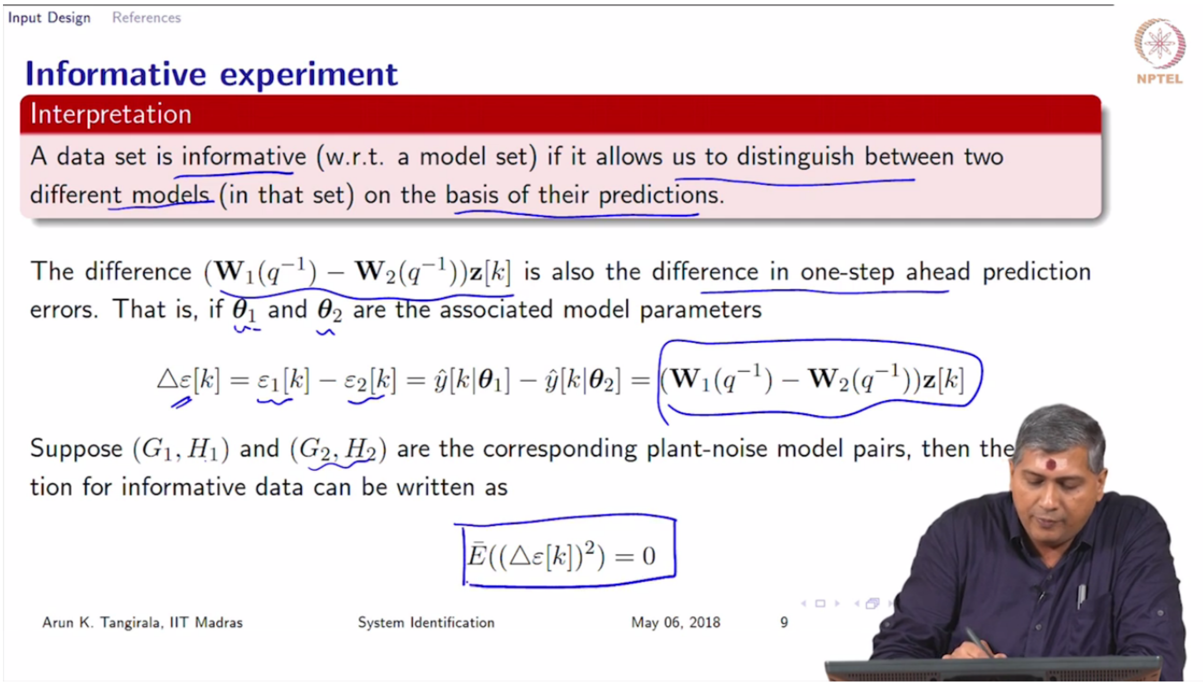
So going back to the interview process, if I placed, if I pick any two candidates I should have asked enough questions to be able to say that this candidate is better suited than the other one. As a simple example, suppose I ask both of them what is your name? One may say, "I'm Ram." Others may say, "I am Sham." Okay? Does that allow me, is there enough information for me to say Ram is better than Sham with respect to the position that I advertised. No, that means now the data is not informative yet. Then I may ask. Can you please explain what is a periodogram?

Then Ram and Sham, if this answer identically, then I have to ask one more question. I have to keep asking until I am able to say, "Yes. One person is technically suited better than the other." But I can't ask arbitrary questions. I have to ask questions relevant to the job there. So here I have to keep the model set in mind, there I have to keep the candidates set in mind. So that is what we mean by informative data. Okay. So in the interview process we go by the answers that they give. In identification we go by the predictions that the model give. That's all.

Now the difference that you have seen earlier W1 minus W2 z is also the difference in the one-step-ahead predictions, you must have noticed that. Therefore if theta 1 and theta 2 are the associated parameters of these two models. Now we are entering the world for para. When we set model set, anyway we are talking of parametric models. So if you do not epsilon 1 as a prediction error from model 1, epsilon 2 as a prediction error from model two, then delta epsilon is a difference between these two. And this is exactly, just to show technically that W1 minus W2 operating on z k is nothing but the difference between the one-step-ahead prediction errors.

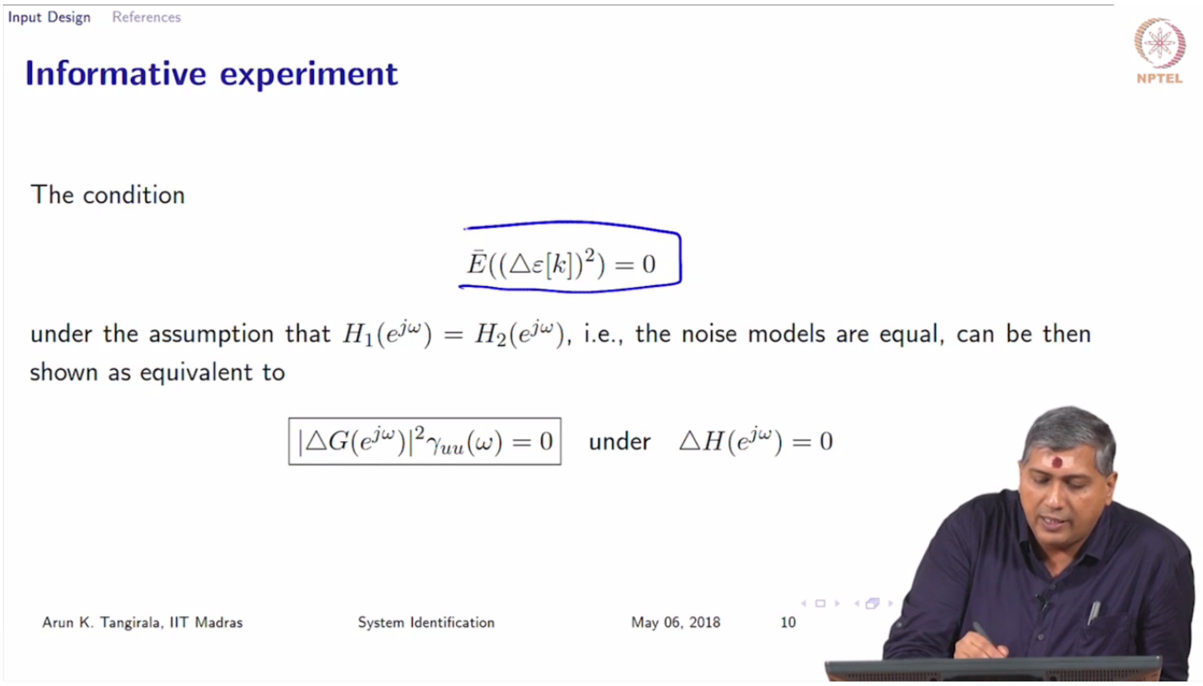
Now suppose G1 H1 is the plant-noise model pair for W1, G2 H25 likewise for W2 then the condition that we have seen for informative data essentially says that if this is, that is.

(Refer Slide Time: 26:44)



This part that you have seen here, this box, here essentially boils down to this. And this should happen. If this is true, let us say it turns out that you pick two models and their prediction errors turned out to be identical. It should happen only when G1 H1 and G2 H2 are identical. If this prediction errors on, in the average, or in the mean square sense turned out to be identical despite being the fact that the noise model or a plant models are different then something is wrong with respect to the information on the data.

(Refer Slide Time: 27:28)



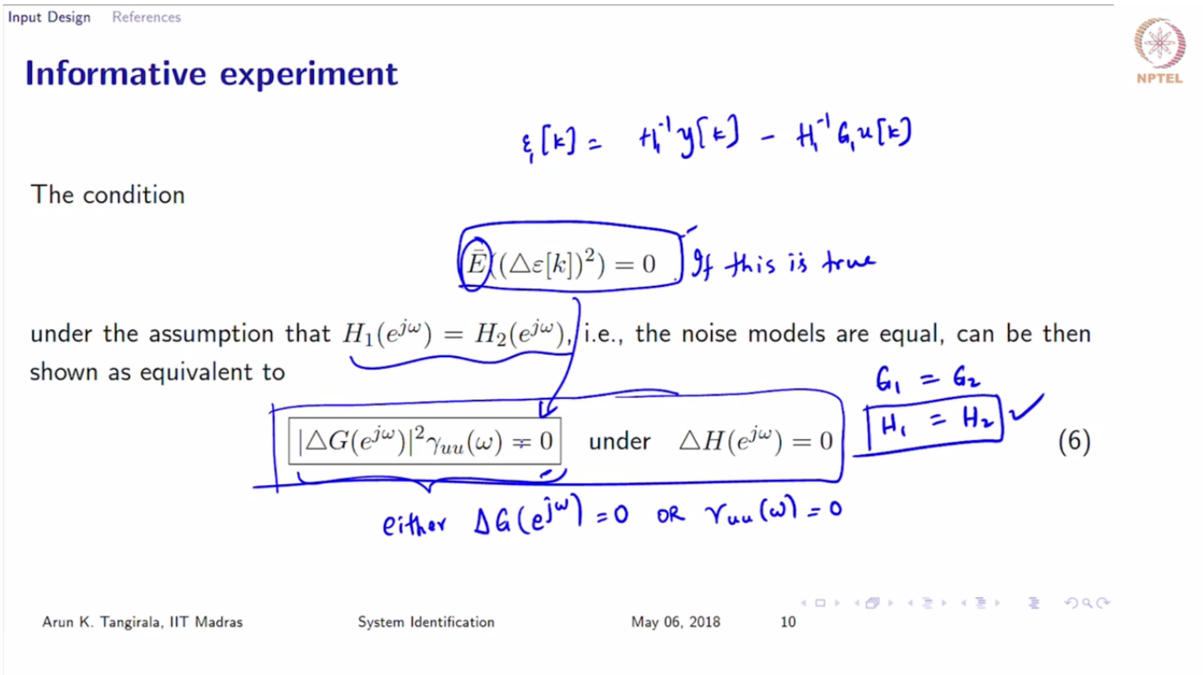
So you can sure that this condition here, using the expression for epsilon under the assumption that the noise models are identical. So we freeze the noise models for now because our focus is on plant models. You can show that if this is true, then this has to be true. Okay. This condition here in the box implies that. How do you start? How do you prove that? Well, essentially you write the expression for prediction error. You recall the prediction error expression. It is y minus y hat. And that is nothing by H inverse y k minus H inverse G u.

You so you can now you write for epsilon 1. You can use H1. And H1 and the [28:29 inaudible] epsilon 2, write the expression for delta epsilon, squared expression. Take the generalised expectation E bar and then you can see the derivation in [28:42 inaudible] or my book as well. And you can show essentially this is the condition. And what did we say, if this is zero, if the prediction error, in the difference in the prediction errors in the mean square hence zero or if this condition is zero, it should happen only when the models are identical.

But this equation tells me that it is also possible that this can be zero when Delta G is not zero. Anyway we have said so, what we mean by what models are identical? G1 should equal G2, H1 should equal H2. We have already frozen, we have said H1 is equal to H2. So what are we demanding now? We are saying that data is informative with respect to a model set if this holds. When can this hold, when can this happen. Either Delta G is zero at a particular frequency or what else is a solution. Gamma u u of omega is 0. One of these two can lead this to zero.

So now very nicely we have started with the equality of models and brought it down to the condition on the spectral density of the input. So simple observation is, this delta G of e to the j omega magnitude squares times gamma equals to 0, can go to 0, even when delta G is not equal to 0. Because gamma u u at that omega may be 0. That means there is no excitation in the input at that frequency. In other words when have not excited the system at a particular frequency, v will not be in a position to say whether one model is better than the other. That is the summary.

(Refer Slide Time: 30:43)



So the main interpretation is that the input spectral density or inference you can say should be such that the equation 6 that we just saw should be true only when two models are identical. And the corollary of that, is that if the input power density is zero at some frequency, then there is no way of saying whether 6 has gone to 0 because delta G has gone to 0 or otherwise. So qualitatively is exactly what I just stated, if I have not excited the system at a particular omega frequency then there is no way, I will be able to say whether one model is better suited than the other at that frequency. At some other frequency maybe but not at that frequency.

(Refer Slide Time: 31:31)